Posted on Authorea 18 Jun 2020 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.159250147.77419720 | This a preprint and has not been peer reviewed. Data may be preliminary.

Türkiye COVID-19 ile mücadelede ne durumda? İstatistiki bir inceleme.

Mert Taşdemir[1]

[1]Affiliation not available

June 18, 2020

Aylardır COVID19 ile yatıp kalkıyoruz. Her gün televizyondan, Twitter'dan sağlık bakanımızın yayınladığı verileri görüyor ve inceleyip kendimizce yorumluyoruz. Ancak bu veriler, aynı zamanda temel istatistik bilgilerimizi gözden geçirmek, matematiksel modellerin nasıl kurulduğuna bir bakış atmak ve bilimin neden önemli olduğunu daha iyi anlamak için ilginç ve öğretici bir fırsat da sunuyor!

Bu yazıda 27 Mart tarihindeki verilerden 17 Hazirana kadar olan istatistikler, MATLAB kullanılarak işlenmiştir. Ancak şuna dikkatinizi çekeriz: Burada veriler genel geçer kabul gören epidemiyolojik modeller çerçevesinde incelenmemektedir ve sadece bir "veri seti" olarak görülmekte, bunun üzerine sanki hiç epidemiyolojik modelimiz yokmuş da sıfırdan bir model geliştirebilir miyiz, bunu sorguluyormuşuz gibi ele alınmaktadır. Bu konuya sonuç kısmında daha detaylı döneceğiz; fakat internette görebileceğiniz "salgın modellerinin" birçoğunun akademik temelden yoksun olduğunu ve bu tarz uyarılar konmaksızın paylaşıldığını hatırlatmak isteriz.

Tekrar etmek gerekirse burada göreceğiniz, epidemiyolojik modeller kullanılmaksızın, bir veri setinin çeşitli istatistiksel analizlerle incelenip, sıfırdan bir model oluşturulması örneğidir. Bu model sınırlı olmak zorundadır; çünkü epidemiyolojik faktörler (örneğin bireyler arası etkileşimler, virüsün bulaşma dinamikleri, vb.) göz ardı edilmekte, ham verilere odaklanılmaktadır. Üstelik sadece Türkiye verileri kullanılmaktadır - ki tek bir ülkenin dinamiklerinin tüm Dünya'daki salgın modellerini yansıtması mümkün değildir. Bu nedenle bunu öğretici bir pratik olarak görmeli, gerçek bir epidemiyolojik model olarak yorumlamamalısınız.

Örneğin bu analizi yaptığımız sıralarda ülkelerin birçoğu katı karantina ve sosyal mesafelendirme önlemlerini gevşetmekteler. Bu durum, incelediğimiz tarih aralığındaki gidişatın bu noktadan sonra tamamen değişmesi anlamına gelmektedir. Bu durumda, buradaki analizler çerçevesinde geçmişe bakarak gelecekle ilgili tahminde bulunmak imkansız olacaktır; çünkü ülkelerin salgına verdikleri tepki değişmekte, bu da dinamikleri alt üst etmektedir. Bu nedenle internette gördüğünüz gelecek tahminlerine çok temkinli yaklaşmanız ve alt yapısını sorgulamanız gerekmektedir.

Öte yandan elbette bu, ham istatistiki analizlerin bize salgınla ilgili hiçbir şey öğretemeyeceği anlamına gelmiyor. En basitinden, denk gelebileceğiniz birçok analizin ne kadar teknik gibi gözüküp de, ne kadar yalın gerçekleri atladığını fark etmenize yarayabilir. Gelin öncelikle, yazıda kullanılacak bazı kavramlardan bahsedelim.

Örneklem Ortalaması, Standart Sapma ve Varyans

Elimizde X_1,..,X_nX1,..,Xn veri grubu olsun:

- Bu veri grubunun örneklem ortalaması$\overline{X} = \frac{\Sigma_{i=1}^{n} X_i}{n}$ ile verilir.
- Bu veri grubunun örneklem varyansı $S^2 = \frac{1}{n-1}\Sigma_{i=1}^{n}(\overline{X} - X_i)^2$ ile verilir.
- Öneklem standart sapması ise S ile verilir; yani varyansın kareköküdür.

Posted on Authorea 18 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159250147.77419720 — This a preprint and has not been peer reviewed. Data may be preliminary.

Not: İstatistikte örneklem ortalaması ve varyansı gibi bilgilerden bütün popülasyonun ortalaması ile varyansını bir miktar hata payı ile tahmin etmeye çalışırız. Bu yazıda da yapacağımız şey aslında bu.

Normal Dağılım

Üniversitede ders gören okuyucular bu dağılıma **çan eğrisi** ismiyle aşinadır. Doğadaki birçok şeyin modellenmesinde kullanılır. Ortalaması μ varyansı σ^2 olan normal dağılım $N(\mu, \sigma^2)$ile gösterilir. Normal dağılımın grafiği aşağıdaki görseldeki gibidir:

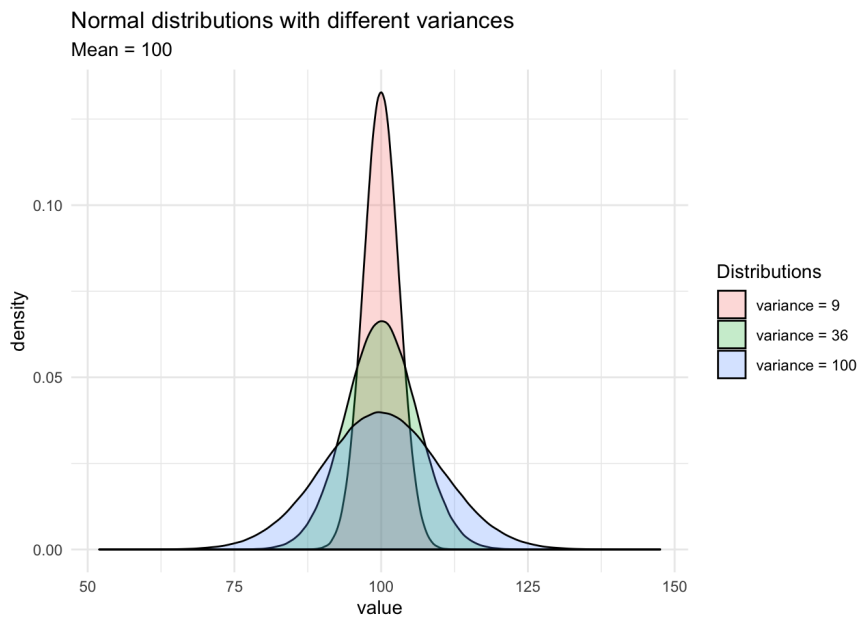

Figure 1: Farklı Varyanslarda Normal Dağılım

Varyans, bize verilerin ortalamadan ne kadar uzaklaşabileceği hakkında bilgi verir. Çoğunlukla popülasyon ortalaması ve varyansını bilmeyiz, öğrenmek için ise aşağıda görebileceğiniz **t-dağılımı** kullanılır. Bu dağılım, normal dağılım **ki-kare dağılımı**nın karekökü arasındaki bir orandan oluşan dağılımdır. Grafiği, yine çan eğrisine benzerdir. Detaylı okuma için ileri okumalarda güzel kaynaklar bulabilirsiniz. İstatistik hakkındaki bu teorik arka planı verdikten sonra, gerçek veriler üzerinden gidebiliriz.

Test Sayılarının Analizi

27 Mart-17 Haziran arası test sayıları şu şekilde: 7533,7641,9982,11535,15422,12496,18757,16160,19664,20065,21400,20023,2
30177,20143,29230,43498,42004,41431,36318,24001,35771,33283,30303,30395, 33687,35605,36187,32722,37351,33332,34821,38
42032,46800,52901

Test sayılarının ortalaması $\overline{X} = 32912$, standart sapması $S = 10534$ olarak hesaplanır.Bu yüksek standart sapma değeri, test sayısını önceden kestirmenin çok zor olduğunu, çünkü her gün test sayısının çok fazla miktarda artıp azaldığını gösteriyor.

Bu nedenle, test sayısının normal dağılıma sahip olup olmadığına bakalım. Bunun için QQPLOT denilen bir test uygulayacağız. MATLAB içerisinde hazır paket olarak gelen "qqplot(T)" komutu, bu testi hemen yapmamızı sağlıyor. Bu testi yapma amacımız, dağılımın normal olup olmadığını öğrenmek.

Posted on Authorea 18 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159250147.77419720 — This a preprint and has not been peer reviewed. Data may be preliminary.

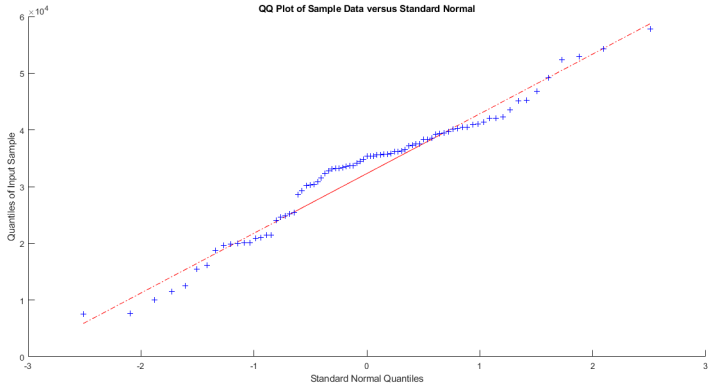

Figure 2: Test sayısı QQPLOT

QQPLOT'tan görülebileceği üzere veriler, normal eğrisine oldukça yakın; dolayısı ile test sayısı istatistiği normal dağılıyor diyebiliriz.

Gerçek ortalamayı şu anda tahmin edebiliriz. Bunu yapmak için, "%98 Güven aralığı" oluşturabiliriz; yani öyle bir aralık oluşturacağız ki, test ortalaması, %98 ihtimalle bu aralığın içindeki bir sayı olacak. Bu aralığı t-dağılımını kullanarak (29871,35953) olarak buluruz. Gerçekten de uygulanan testlerin az önce bulduğumuz ortalamasının da bu aralıkta olduğuna, hatta güven aralığımızın tam ortasında olduğuna dikkatinizi çekeriz. Bu tesadüf değil; çünkü %98 katsayılı güven aralığı $(\overline{X} - \frac{t_{87}^{0.01}S}{\sqrt{(83)}}, \overline{X} + \frac{t_{87}^{0.01}S}{\sqrt{(83)}}))$ burada $t_{87}^{0.01}$ sayısı, t- dağılımına sahip t^* rassal değişkeni için $P(t > t_{87}^{0.01}) = 0.01$ koşulunu sağlayan bir sayıdır.

Salgının sonuna vardığımızda göreceğiz ki Türkiye'nin günlük test ortalaması neredeyse kesin bir ihtimalle bu aralıkta olacak. Aynı aralığı standart sapma için oluşturursak, (773.22,1554.1) aralığını buluruz; yani pek de beklenen bir sapma değil çünkü şu anda hiç de bu kadar az sapmıyor. Bu, bize şunu söyler: Bütün testler bittiğinde, test sayısının standart sapması gittikçe azalacak ve bu aralığa düşecek. Test sayılarını grafiğe dökmek istersek, şöyle bir sonuç elde ederiz:

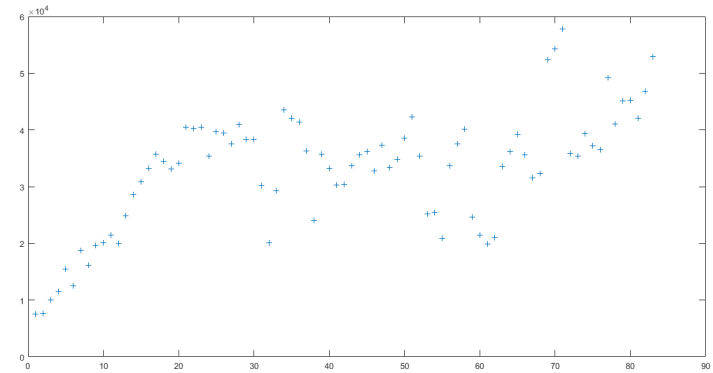

Figure 3: Test Sayısı-Gün grafiği

İstikrarlı bir yükseliş ve sonra verilerin bir denge noktasına gelme çabası rahatça grafikte görülüyor.

Posted on Authorea 18 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159250147.77419720 — This a preprint and has not been peer reviewed. Data may be preliminary.

Vaka Sayısı Analizi

Test sayısının analizini iyice yaptık, şimdi günden güne vaka sayılarına bakalım. 27 Mart- 17 Haziran arası vaka sayıları şu şekilde:

2069,1704,1915,1610,2704,2148,2456,2786,3013,3135,3148,3892,4117,4056,4747,5138,4789, 4093,4062,4281,4801,4353,3783,397

Bu aralıkta açıklanan 83 verinin ortalaması, $\overline{X} = 2159$, standart sapma $S = 1236.9$ olarak hesaplanır. Bu veri de ilginç şeyler söylüyor: Örneğin ortalamaya göre çok yüksek bir sapmaya sahiptir, bu da bir sonraki günkü vakayı kestirmekte zorlanacağımızı söyler. Bu nedenle de "gelecek tahmini" yapan algoritmalara halen güvenmek mümkün değildir.

Standart sapma için %98 güven aralığı oluşturursak, (90.79,182.48) aralığına varırız, yani bu aslında yeni vaka sayısının sapmasının aslında azalacağı anlamına gelir. Tabii yeni vaka sayısı yüksek sayılarda ise bu sonuç kötü bir sonuçtur; ancak gün-yeni vaka sayısının grafiğine bakarsak:

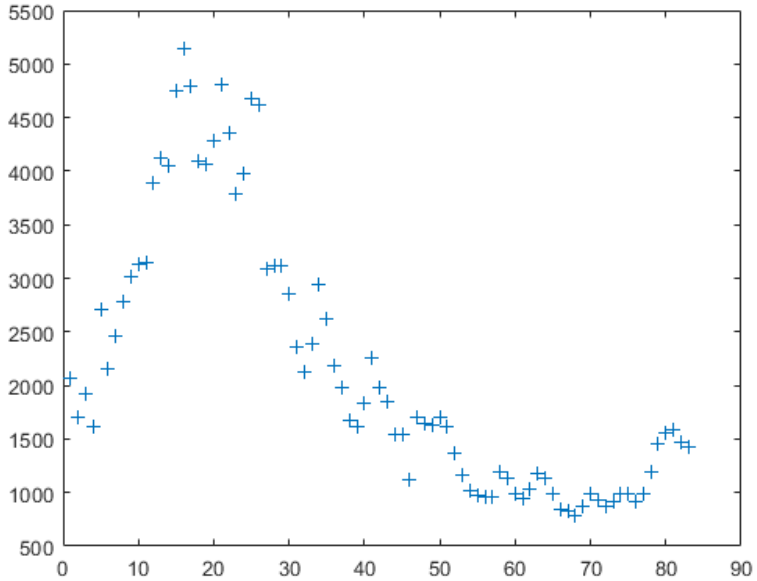

Figure 4: Gün-Vaka sayısı

Günlük yeni vaka sayısının azaldığından ve sapma için verilen güven aralığından, eğer her şey aynı kalırsa, günlük vaka sayısının kolay kolay artmayacağı görülebilir. Ancak unutmamak gerekiyor ki ülkelerin "normale dönme çabaları" gibi çabalar, koşulların aynı kalmamasına neden olmaktadır; dolayısıyla bu tür temel analizlerin hatalı sonuçlar vermesine neden olabilir.

Ölü Sayısı Analizi

67 günlük bu süreçte toplam ölü sayısı 4488, günlük ortalama ölü sayısı $\overline{X} = 57.66$ standart sapma S=35.83. Görülebileceği üzere son zamanlarda açıklanan ölü sayısından çok daha fazla bir sayı bu. Son 30 günlük süreçteki ölü ortalaması 23.25,standart sapması ise 5.72 olarak bulunur, bu çok daha günümüze yakın bir istatistiktir. Zaten eğer bir normalleşme sürecinden bahsediyorsak böyle bir tabloyla karşılaşmamız gerekiyordu. Sapmanın ve ortalamanın gittikçe düştüğünü görüyoruz.

83 günlük süreç için ortalama ve sapmanın güven aralıkları sırasıyla (47.31 ,68) ve (2.63,5.28) olduğunu görürüz. Yani son bir aydaki düşük ortalama güven aralığının dışında kalır, bu da hızlı bir düşüş olduğu anlamına gelir. Standart sapmada ise düşüşte yola devam anlamı çıkar. Ölüm sayıları için qqplot yaptığımızda verilerin normal dağılmadığını görüyoruz.

Tabloda günlük ölü sayılarına baktığımızda ise aşağıdaki gibi bir tabloyla karşılaşıyoruz.

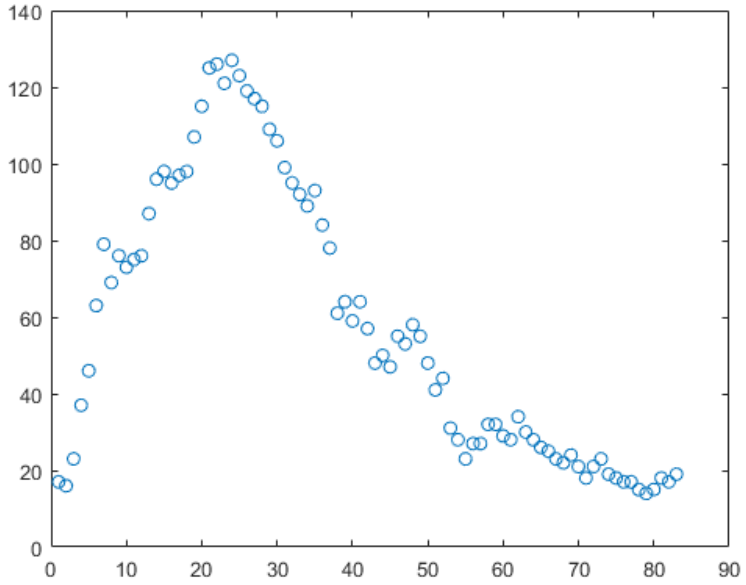

Figure 5: Gün-Vefat

Ölü sayısının normal dağılımdaki gibi bir çan eğrisine sahip olmadığını burada da görebiliriz ancak birkaç cebirsel numara ile normal dağılıma benzetebiliriz. Yani hiçbir şey değişmeseydi, ölü sayısının yavaş yavaş azalmasını beklerdik.

İyileşme Sayısı Analizi

Veriler yayınlanmaya başladıktan birkaç gün sonra yayınlanmaya başlayan iyileşme sayıları verilerine bir göz atalım şimdi. 83 günlük süreçte başlangıçta hiçbir iyileşen açıklanmadı. Bu, elbette kimsenin iyileşmediği anlamına gelmiyor; sadece iyileşen sayılarının açıklanmadığı anlamına geliyor. Bu nedenle ortalama ve standart sapma hesaplarına bu kısmı katmayacağız. 83 günlük süreçteki günlük iyileşen kişi sayıları şu şekilde:

302,256,284,256,264,296,281,542,481,511,842,875,1415,1542, 1822,1523,1454,1488,1559,2014,3246,3845,3558,4651,5018,5231,

Verilere bakıldığında, toplam 154640 kişi iyileşmiş, günde ortalama 1857 kişi iyileşiyor ve günde iyileşen kişi sayısının sapması 1529.4. Yüksek bir sapma, salgın düşüşe geçiş eğiliminde olduğundan avantajlı bir faktör; çünkü düşüş eğiliminde bir sapmanın varlığı, iyileşen sayısında artış ihtimalini arttıracaktır. Güven aralıklarına baktığımızda ise ortalama ve standart sapma için güven aralıkları sırasıyla (1415.8,2298.8) ve (112.26, 225.63) olur.

İyileşen Sayısı - qqplotMATLAB

Posted on Authorea 18 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159250147.77419720 — This a preprint and has not been peer reviewed. Data may be preliminary.

Posted on Authorea 18 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159250147.77419720 — This a preprint and has not been peer reviewed. Data may be preliminary.

Yukarıdaki tabloya bakıldığında verilerin normal dağılmadığı görülüyor. Normal dağılım olmaması, aslında veri sayısının yeterli boyuta ulaşmadığını, yani günlük iyileşen sayısının daha da artabileceğini gösteren, önemli bir veri.

Av-Avcı Dinamiği: Tüm Bunlardan Bir Model Geliştirmek

Bu verilerden daha net sonuçlar çıkarmak için özellikle gün-vaka ve gün-iyileşen hasta grafiklerini bir arada görüp karşılaştırmakta fayda var.

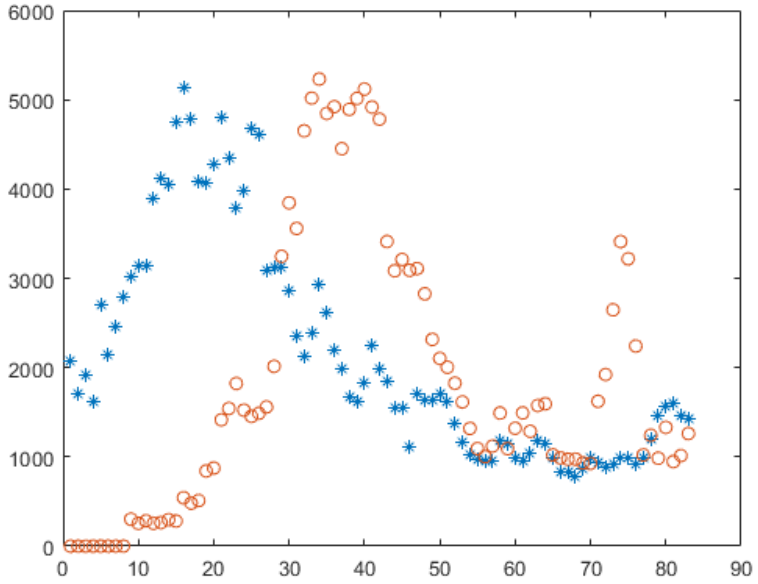

Figure 6: Vaka-İyileşen sayısı grafikleri bir arada.

Burada içi boş daireler ile işaretlenen veriler günlük iyileşen kişi sayısını, yıldızlar ise günlük vaka sayısını temsil etmektedir. Vaka sayısının zirveye ulaştığında iyileşme sayısı diplerde iken, iyileşme sayısı zirvede olduğunda vaka sayısının dibi gördüğünü görüyoruz. Bu, aslında bir av-avcı ilişkisine benzetilebilir.

Posted on Authorea 18 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159250147.77419720 — This a preprint and has not been peer reviewed. Data may be preliminary.

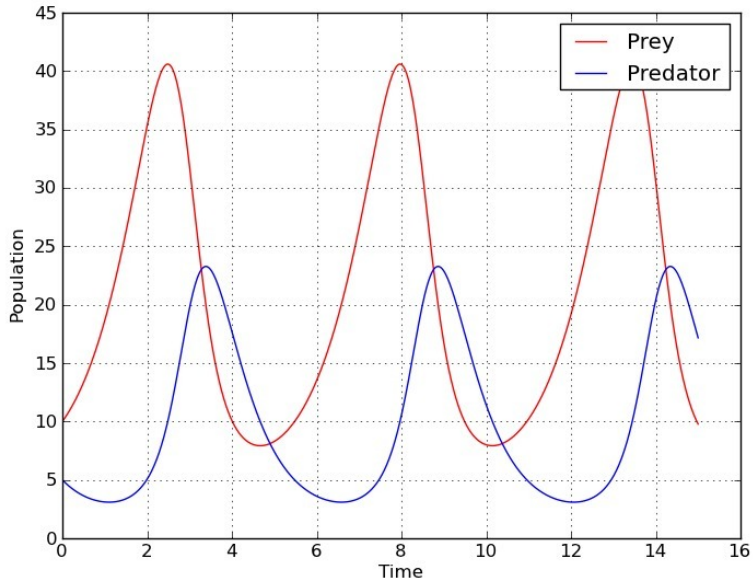

Figure 7: Av-Avcı grafiği

Burada kırmızı renkli bölüm av olan hayvanın popülasyonunu gösterirken, avcı popülasyonu mavi ile işaretlenmiştir. Av-avcı arasını ifade eden diferansiyel denklem sistemi

$\frac{d\lambda}{dt} = a\lambda + b\lambda\mu$

$\frac{d\mu}{dt} = c\mu + dlm$

ile modellenir ve bu eşitliklere aynı zamanda **Lotka-Volterra denklemi** denir. Burada a, b, c, d sayıları av ile avcı arasındaki ilişkilerle belirlenen parametrelerdir;av ile avcının doğum ve ölüm oranları ile ilgilidir. t zamanı belirtir ,λ,μ sırasıyla av sayısı ile avcı sayısını belirtir. Bu grafiklerdeki benzeşmeler, bizi günde iyileşen sayısı ile günlük vaka sayısı arasında bir modelleme yapmaya teşvik edebilir.

Lotka-Volterra denklemini bu modele uyarlayalım.

λ ,μ sırasıyla günlük vaka sayısı ile günlük iyileşen sayısı olsun. Bu diferansiyel denklem sistemini çözersek $d\lambda + c\ln\lambda + b\mu + a\ln\mu = \zeta$ (burada ζ başlangıç koşullarından elde edilen bir katsayıdır) sonucuna varılır.

4 Nisan tarihini başlangıç olarak alırsak (çünkü ilk iyileşen sayısı bu gün yayınlandı) ,λ(4.4.2020)=3013 ve μ(4.4.2020)=302 başlangıç değerlerine ulaşırız. Bu başlangıç değerlerinden ζ 'yı a,b,c,d cinsinden bulabiliriz. Hesap yapınca 3013d+8.01c+302b+5.71a=ζ sonucuna varırız o halde derin bir istatistiksel araştırma yaptıktan sonra iyileşen/günlük vara arasında a,b,c,d'yi de bulabilir ve modeli dλ+clnλ+bμ+alnμ=3013d+8.01c+302b+5.71a olarak buluruz.

Sonuç ve Uyarı

Bütün bu yapılan istatistiki işlerden yola çıkılarak varılabilecek sonuçların en önemlisi salgının yavaş yavaş dengeye gelmekte olduğu ancak dengeye gelmenin salgının bitmesi anlamına gelmediğidir. Yukarıda yapılan bütün istatistiklerin bir anda değişmesi(Vaka sayısında ortalama artışı gibi) bizlere bağlı. Bu istatistiklerden

geleceğe yönelik yapılan tahminler keskin tahminler değildir çünkü basit istatitiki yöntemler kullandık. Bu yazıdaki en temel amacımız salgının şu ana kadarki gidişat hakkında formal yorumlarda bulunmaktır. Tabii sonda bir de süpriz uygulama bölümü var.

Bulduğumuz bu model tabii ki yaklaşık bir modeldir ve çok daha kesin sonuçlar veren epidemik modeller matematikçiler tarafından bulunmuştur. Bu yazıda başta da söylendiği gibi önce istatistik hakkında teorik bir ön bilgi verip daha sonra Sağlık Bakanlığının yayınladığı koronavirüs istatistikleri ile basit bir istatistik çalışması yaptık ve bu alana uzak gibi görünen ancak oldukça işimize yarayan av avcı denklemlerinden bir model ürettik. Konular ilginizi çektiyse ileri okumalar bölümünden araştırmaya devam edebilirsiniz, bilimle ve sağlıkla kalınız.

Posted on Authorea 18 Jun 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159250147.77419720 — This a preprint and has not been peer reviewed. Data may be preliminary.